Effects of Ability Grouping on Students' Attitudes toward Korean and English

Seon-Hi Shin and Seohyun Kim
(Seoul National University)

Shin, Seon-Hi and Kim, Seohyun. (2012). Effects of Ability Grouping on Student's Attitudes toward Korean and English. ***Language Research*** **48.3, 693-714.**

The study aimed to investigate the effects of ability grouping in Korean and English on students' academic performances and their attitudes toward subjects. A multiple group analysis with a 2-level HLM was performed on large-scale data using SAS, R, Mplus, and HLM for comparison. The study showed that the effects of ability grouping on student academic achievement and student perception about curricula were ambiguous. The study also showed that Mplus and HLM produced larger standard errors of the parameter estimates than SAS and R.

Keywords: Korean or English language proficiency, ability grouping, hierarchical linear modeling, multiple group analysis

1. Introduction

Student achievement is a national and global concern as shown in recent large scale standardized assessments such as the Trends in International Mathematics and Science Study (TIMSS), the Program for International Student Assessment (PISA), and the National Assessment of Educational Achievement (NAEA). Over the past decade, Korea has ranked at or near the top in international tests (Foy & Olson 2009; OECD Programme for International Student Assessment, 2009). Many studies have explored important factors contributing to student achievement in various content areas including reading literacy. The findings of such studies have inspired or alerted educational stakeholders to change existing school policies or launch new educational initiatives for improving student achievement in schools. As technological and economic competition has intensified globally, academic excellence has become a major focus in education in many countries.

On the other hand, educational equality also continues to draw attention from the education community as part of the central mission of schools.

Academic excellence and educational equality are also being emphasized in language education. Ability grouping is often seen as a policy to commit schools to academic equality and excellence. It allows different ability groups to receive educational instruction and learning opportunities matched to their needs and preparations. For instance, it lays out different curriculum paths for students who wish to go to college and for those who are headed for the workplace. Dividing academic classes geared to different levels for students of different abilities is called ability grouping, which is a form of tracking (Gamoran 1992; Oakes 2000). According to Oakes, the underlying assumptions of ability grouping are (1) that the academic needs of all students will be better met when they learn in groups with similar capabilities or prior levels of achievement, (2) that less capable students will suffer emotional as well as educational damage from daily classroom contact and competition with their brighter peers, (3) that students can be placed in tracks or groups both accurately and fairly, and (4) that most teachers and administrators contend that ability grouping greatly eases the teaching task. However, the findings have been ambiguous. Many studies have reported that more learning occurs in higher tracks (Ball 1981; Barr & Dreeben 1983; Burgess 1984; Gamoran 1987; Gamoran & Mare 1989; Hoffer 1992; Oakes 1985;), and that ability grouping does not increase student learning but amplifies educational inequality (Rosenbaum 1976; Oakes 2000).

Meanwhile, Korea adopted ability grouping relatively recently. Since the 7th Curriculum Reform Act came into effect in 2000, the elementary and secondary schools in Korea have gradually expanded a program of ability grouping on the premise that the growth potential of individual students and the effectiveness of schooling can be maximized when education is provided based on academic needs, interests, aptitudes, and abilities of individual students (Yu J. 2008). According to the 2009 NAEA database (Ministry of Education, Science and Technology, 2010), about 1% of all high schools had an ability grouping policy in all five major subjects, which are Korean language arts (hereafter, Korean), social science, mathematics, science, and English, whereas only 19% had no ability grouping policy in any of the subjects.

Further, 76% of the schools were reported to employ ability grouping in both English and mathematics classes, and the percentages of schools that grouped students by ability in Korean, science, and social science classes were 5%, 1.5%, and 1.4%, respectively.

Regarding the effects of ability grouping in classes, the literature has reported controversial findings (Choi J. 2006; Hwang Y. & Kim G. 2008; Kim H. 2006; Seo H. 2008). For instance, a study reported that ability grouping improved student achievement in the low ability group whereas the degree of achievement decreased in the high ability group (Hwang Y. & Kim G. 2008). Students' perception of ability grouping turned out both positive and negative in the studies. In a study (Seo H. 2008), a majority of students agreed with the desired goals of ability grouping but complained that the way the policy had been implemented favored the high ability group. Researchers also suggested ways to improve the policy (Seo H. 2008; Yu J. 2008).

With regard to ability grouping, it also seems important to inquire whether or not students' attitudes toward the curriculum change from experiencing ability grouping in school. It is because students' perceptions are believed to play a significant role in academic success or delinquency (Ban J. & Shin S. 2011). For instance, a study reported that ability grouping was positively associated with academic commitment, and that the commitment was negatively related to school crime in path analysis (Jenkins 1995).

2. Objectives

The current study aimed to investigate the effects of ability grouping in Korean and English on students' academic performances and their attitudes about subjects. Specifically, the major research questions included (1) whether ability grouping was positively associated with student performance in a government-mandated large-scale test, and (2) how students' attitudes toward curriculum differed between schools where ability grouping had already been implemented and schools with no such policy. Some covariates that were believed to influence students' attitudes were taken into account while comparing the two ability-grouping conditions. The covariates included in the analysis were the ratio of the number of students from homes receiv-

ing welfare money to the total number of students (SES), the number of teachers to the student body or teacher rate (TCHRATE), time spent in doing homework (HW), attending private tutoring classes (EXTR), or taking EBS or online courses (EBS).

SAS, Mplus, R, and HLM software packages were used to perform multilevel modeling for the second question. It is important to know which software is a reliable and accurate analytical tool for parameter estimation in hypothesis testing. Thus, the third objective of the research was (3) to compare results from different modeling software packages, and to explain the factors contributing to the similarities and/or differences.

3. Methods

3.1. Data Sources

This study analyzed achievement and survey data from the National Assessment of Educational Achievement (NAEA) administered to high school students in 2009. All high schools, regardless of the school type such as national, public, or private schools, were included in the analysis. Vocational schools and mixed-purpose schools were also included. The NAEA (Ban J. 2006) is a government-mandated achievement test to assess whether students have acquired content knowledge and performance skills aligned to common educational standards in five content domains, which are Korean, mathematics, science, English, and social studies. The test provides test takers from three grade levels − 6, 9, 11 − with scale scores and performance levels. The goal of NAEA is to examine and improve the quality of school education at the national level. Specifically, the purpose of assessing each subject is to examine students' achievement level in the subject and, based on this information, explore ways to enhance learning of the subject matter. The scale score was developed in 2003 such that the mean should be 260 and the standard deviation 8.5 score points (Ban J. 2006). The scale was maintained through IRT-based equating until 2010 when a new scale was developed.

The students who took the test were asked to fill out a survey questionnaire as well. Their teachers and principals were also surveyed.

For the present study, students' test scores from the 2009 administration were analyzed with their survey responses alongside the principals' survey responses. The principals' responses to the survey items provided information on the characteristics of the school that a student attended whereas the students' survey responses contained their personal background information. After merging the multilevel data files, the cases having missing data on any of the covariates as well as the performance and attitude scores were excluded. The number of student records being used in the analysis was 626,098, and a total of 2,184 school records were included. Appendix A shows the research variables and the survey questions on the basis of which the variables were composited.

3.2. Data Analysis

The NAEA test was administered to all high schools instead of a random sample of schools in 2009. The observations were not independent but correlated with each other: students' perception scores were nested within a certain school. Furthermore, surveys were administered to students and schools' principals to gather the contextual information surrounding schools and students alike. Hence, the data was multilevel with students at the lower level and schools at the upper level. When a conventional multiple regression is applied to multilevel data, the Type-I error rate associated with a significance test is known to be inflated (Bryk & Raudenbush 1992; Littell, Milliken, Stroup, & Wolfinger 1996; Verbeke & Molenberghs 2000).

Therefore, when the dependent variable was student perception about curriculum (YCURR), the data was analyzed by formulating a 2-level hierarchical linear model (HLM). Additionally, since the study aimed to compare two ability-grouping conditions in each subject − i.e., having an ability grouping policy or not having such a policy in Korean or English −, a group membership variable had to be incorporated into the model for the sake of direct comparison. This variable, i.e. GRP_KOR or GRP_ENG, was coded such that schools having an ability-grouping policy in the subject were assigned 1 whereas those with none were assigned 0. Hence, the method can be called a multiple group analysis using a 2-level HLM. Models were built separately for these two languages. In the HLM the attitude

score about each subject was used as the criterion variable, and ability grouping and covariates at the student- and school-levels were predictor variables whose associations with the criterion variable were investigated.

For example, the model with GRP_KOR was formulated only with the ability grouping variable first. Then, the other predictors were added to the model at student- and school-levels simultaneously. In multilevel modeling, the former is called the unconditional model while the latter is the conditional model. The unconditional model for the study was parameterized at the student level as follows:

$$(\text{YCURR})_{ij} = \pi_{0j} + e_{ij}, \quad (1)$$

where $(\text{YCURR})_{ij}$ was an attitude score for student i in school j, and the expected attitude score across students for school j was denoted by π_{0j}. On the other hand, e_{ij} indicated a random error associated with each student. The random error was assumed to be normally distributed with a mean 0 and a variance σ^2. Subsequently, the level-2 model was specified

$$\pi_{0j} = \beta_{00} + d_{0j}, \quad (2)$$

where β_{00} was the expected attitude score across the two ability-grouping conditions while d_{0j} denoted a random school deviation from the expected value. The deviation was also assumed to be normally distributed with a mean of 0 and a variance of τ^2. Then, the β_{00} was further decomposed to compare the two conditions directly such that

$$\beta_{00} = \gamma_{000} + \gamma_{001} (\text{GRP\_KOR})_j, \quad (3)$$

where the dummy variable GRP_KOR took on a value of either one or zero to reflect the ability-grouping policy of the school. As a result, the unconditional HLM for the multiple group analysis of the current study was

$$(\text{YCURR})_{ij} = \gamma_{000} + \gamma_{001}(\text{GRP\_KOR})_j + \{d_{0j} + e_{ij}\}. \quad (4)$$

The model had two fixed effects along with the two random effects

within the curly bracket. The symbol γ_{000} represents the expected attitude score for schools having no ability-grouping policy and γ_{001} indicates the difference of the expected attitude score for the schools having an ability-grouping policy from that of the schools without such a policy. The variances of the random effects were σ^2 and τ^2, respectively, and the random effects were independent of each other.

The extension of the HLM into the conditional HLM was straight forward. First, the level-1 model had the following predictors:

$$(\text{YCURR})_{ij} = \pi_{0j} + \pi_{1j}(HW)_{ij} + \pi_{2j}(EXTR)_{ij} + \pi_{3j}(EBS)_{ij} + e_{ij}. \quad (5)$$

With two school background variables resulting from principals' survey responses, the level-2 model was specified as

$$\begin{aligned} \pi_{0j} &= \beta_{00} + \beta_{01}(SES)_j + \beta_{02}(TCHRATE)_j + d_{0j}, \\ \pi_{1j} &= \beta_{10}, \\ \pi_{2j} &= \beta_{20}, \\ \pi_{3j} &= \beta_{30}. \end{aligned} \quad (6)$$

where the school background variables were included as predictors for intercept π_{0j} only. The effects of homework, extracurricular tutoring, and EBS classes were assumed to be fixed across schools. Then, each beta was further decomposed to represent the two ability-grouping conditions as shown below.

$$\begin{aligned} \beta_{00} &= \gamma_{000} + \gamma_{001}(\text{GRP\_KOR})_j \\ \beta_{01} &= \gamma_{010} + \gamma_{011}(\text{GRP\_KOR})_j \\ \beta_{02} &= \gamma_{020} + \gamma_{021}(\text{GRP\_KOR})_j \\ \beta_{10} &= \gamma_{100} + \gamma_{101}(\text{GRP\_KOR})_j \\ \beta_{20} &= \gamma_{200} + \gamma_{201}(\text{GRP\_KOR})_j \\ \beta_{30} &= \gamma_{300} + \gamma_{301}(\text{GRP\_KOR})_j \ . \end{aligned} \quad (7)$$

The school deviation from the group mean of the intercept $-d_{0j}-$ was assumed to be normally distributed with a mean of 0 and a variance of τ^2. The random error e_{ij} was also assumed to follow a normal distribution with a mean of 0 and a variance of σ^2. All the variance components were assumed to be independent of each other in this model.

As shown in the model formulation, the ability-grouping variable did not constitute a Level-3 above the school level because the two conditions or groups were not randomly drawn. They were specifically categorized with the intention for direct comparison, and the categories were exclusive and exhaustive. The combined conditional HLM was

$$\begin{aligned}(\text{YCURR})_{ij} = \{&\gamma_{000} + \gamma_{001}\,(\text{GRP\_KOR})_j + \gamma_{010}(SES)_j + \\ &\gamma_{011}(\text{GRP\_KOR})_j *(SES)_j + \gamma_{020}(TCHRATE)_j + \\ &\gamma_{021}(\text{GRP\_KOR})_j *(TCHRATE)_j + \gamma_{100}(HW)_{ij} + \\ &\gamma_{101}(\text{GRP\_KOR})_j *(HW)_{ij} + \gamma_{200}(EXTR)_{ij} + \\ &\gamma_{201}(\text{GRP\_KOR})_j *(EXTR)_{ij} + \gamma_{300}(EBS)_{ij} + \\ &\gamma_{301}(\text{GRP\_KOR})_j *(EBS)_{ij}\} + \{d_{0j} + e_{ij}\}, \end{aligned} \tag{8}$$

where the first and second curly brackets contained the fixed and random effects of the model, respectively. The components of the total variance were σ^2 at the student level, and τ^2 at the school level. The random effects were independent of each other. The regression coefficient γ_{000} denoted the expected attitude score for the schools having no ability-grouping policy when all the covariates and their interaction terms were zero.

4. Results

As shown in Tables 1 and 2, an overwhelming majority of schools (95%) had no ability-grouping policy for Korean whereas only 12.5% of schools had no ability-grouping policy for English. Looking at the column percentages in the tables, schools having no ability-grouping policy for Korean performed slightly better in the Korean test. In contrast, schools having ability-grouping policy for English appeared to perform better in the English test. The percentage of students attaining to the Proficient level or higher in Korean was 78% for schools without an ability-grouping policy as compared to 74% for schools with one (Table 1). On the other hand, the percentages in English were 36% for schools without an ability-grouping policy and 57% for schools with one (Table 2).

The association between ability grouping and performance was fur-

ther tested by constructing a hierarchical generalized linear model (HGLM). The schools having an ability-grouping policy for Korean performed significantly lower at the .05 significance level ($t(2182) = -1.98$). The schools having an ability-grouping policy for English showed a significantly higher level at the same significance level ($t(2182) = 9.877$). The analysis was performed using the commercial statistical software HLM.

Table1. Frequencies of Performance Level by Ability-Grouping Policy in Korean

Performance level	Ability-grouping policy	
	Absent (%)	Present (%)
Advanced	191,193 (31.98)	8,028 (28.42)
Proficient	275,284 (46.05)	288,098 (45.37)
Basic	108,896 (18.21)	6,009 (21.27)
Below basic	22,480 (3.76)	1,394 (4.94)
Total	597,853 (95.49)	28,245 (4.51)

Note. The number within each pair of parentheses represents the column percentage.

Table 2. Frequencies of Performance Level by Ability-Grouping Policy in English

Performance level	Ability-grouping policy	
	Absent (%)	Present (%)
Advanced	14,030 (17.92)	89,356 (16.31)
Proficient	14,304 (18.27)	219,380 (40.50)
Basic	38,121 (48.70)	205,755 (37.56)
Below basic	11,827 (15.11)	33,325 (6.08)
Total	78,282 (12.50)	547,816 (87.50)

Note. The number within each pair of parentheses represents the column percentage.

The pairwise correlations among the predictors are shown in Table 3. The absolute magnitude of the correlations ranged from .04 to .44. As a rule of thumb, the predictors are considered as having little to low correlation with each other (Stevens 2002). The correlations of GRP_

KOR with the other predictors were much weaker ranging from -.00 to -.01. Hence, multicollinearity may not be a concern with the current data.

Table 3. Correlation Coefficients for Predictors for English

	(1)	(2)	(3)	(4)	(5)	(6)
(1) SES						
(2) TCHRATE	.437					
(3) HW	-.132	-.046				
(4) EXTR	-.320	-.208	.203			
(5) EBS	-.158	-.096	.267	.294		
(6) GRP_ENG	-.263	-.244	.036	.133	.088	

SAS (Version 9.3), R (Version 2.15.1), Mplus (Version 6.12), and HLM (Version 6.01) were used to fit HLMs for the study. Missing data were deleted list wise, meaning that complete cases were included in the analysis. The estimates of the fixed effects and the variance components of the unconditional model for attitude toward curriculum are shown in Tables 4 and 5. The expected score of the schools having no ability-grouping policy (γ_{000}) was 20.03 for Korean and 19.65 for English uniformly across the software. The attitude toward curriculum for the schools having an ability-grouping policy for Korean was slightly less positive. The amount of the change between the two groups was -0.21 (γ_{001}) As for English, the amount of the change between the two groups was 0.468, and it was statistically signifiant.

The two independent variance components at the two levels of the model were estimated to capture the variability existing between students (σ^2) and between schools (τ^2). The between-students variance, for example, reflected the extent to which the students' scores varied randomly from the school mean while the between-schools variance expressed the degree to which the schools' scores deviated randomly from the ability-grouping condition's mean. The proportion of the between-schools variance in the total variance is called the intra-class correlation (ρ) in the literature (Bryk & Raudenbush 1992). As the intra-class correlation increases, the difference among schools becomes

larger. The intra-class correlation was around .10 for both Korean and English, suggesting that the schools were relatively homogeneous in the attitude score. All the statistics for Korean and English remained quite similar across the four software packages.

Table 4. Parameter Estimates of the Unconditional Model for Korean

Fixed effect coefficient (SE)	Software			
	SAS	R	Mplus	HLM
Intercept, γ_{000}	20.034* (0.026)	20.034* (0.026)	20.034* (0.025)	20.034* (0.026)
GRP_ENG, γ_{001}	-0.205 (0.113)	-0.205 (0.113)	-0.203 (0.115)	-0.205 (0.115)
Random effect				
Intra-class correlation, ρ	0.099	0.099	0.099	0.099
Level-1 variance, σ^2	11.600	11.600	11.600	11.600
Level-2 variance, τ^2	1.269	1.269	1.269	1.270

* $p < .05$

Table 5. Parameter Estimates of the Unconditional Model for English

Fixed effect coefficient (SE)	Software			
	SAS	R	Mplus	HLM
Intercept, γ_{000}	19.645* (0.056)	19.645* (0.056)	19.645* (0.081)	19.645* (0.081)
GRP_KOR, γ_{001}	0.468* (0.063)	0.468* (0.063)	0.468* (0.084)	0.468* (0.085)
Random effect				
Intra-class correlation, ρ	0.097	0.096	0.096	0.096
Level-1 variance, σ^2	11.600	11.600	11.600	11.600
Level-2 variance, τ^2	1.239	1.240	1.240	1.240

* $p < .05$

When the conditional model was fit, the percentage of the between-students variance (σ^2) explained by the three level-1 predictors (HW, EXTR, and EBS) was 12% for both models with ability grouping in

Korean and English. The reduction in the variance at the school-level by the predictors (SES, TCHRATE) was about 66% for both models (Tables 6 and 7).

Regarding the fixed effects for the model of attitude toward curriculum, the schools having an ability-grouping policy for Korean showed significantly less positive attitude with all the covariates taken into account (γ_{001}). The number of students from homes receiving welfare money relative to the total number of students was negatively associated with attitude toward curriculum (γ_{010}). This negative relationship was amplified somewhat, but not significantly more obvious for schools having no ability-grouping policy (γ_{011}). The ratio of the number of teachers to the total number of students (γ_{020}), the time spent in doing homework (γ_{100}), private tutoring classes (γ_{200}), or EBS online courses (γ_{300}) were all positively related with attitude toward curriculum. These positive relationships were significantly more salient for the schools having an ability-grouping policy in the results of Mplus and HLM. All the statistics turned out quite similar across the four software packages (Table 6).

The finding from fitting the model for English showed that the schools having an ability-grouping policy were less positive in attitude about curriculum (γ_{001}). However, the difference was not statistically significant. As the number of students from welfare families increased relative to the total number of students, attitude toward curriculum became negative (γ_{010}). This negative effect was significantly less evident for schools having an ability-grouping policy (γ_{011}). The ratio of the number of teachers to the total number of students (γ_{020}), the time spent in doing homework (γ_{100}) or EBS online courses (γ_{300}) were all positively related with attitude toward the subject. The positive effects of the teacher rate and taking EBS online courses were significantly reduced with the schools having an ability-grouping policy. The positive effect of the time spent in private tutoring classes was significantly intensified with the schools having an ability-grouping policy. Overall, all the statistics remained similar across the four software packages (Tables 7). SAS and R yielded more similar results while Mplus and HLM produced more similar estimates.

Table 6. Parameter Estimates of the Conditional Model for Korean

Fixed effect coefficient (SE)	Software			
	SAS	R	Mplus	HLM
Intercept, γ_{000}	16.268* (0.040)	16.268* (0.040)	16.269* (0.064)	16.268* (0.063)
GRP_KOR, γ_{001}	-0.896* (0.225)	-0.896* (0.225)	-0.896* (0.343)	-0.895* (0.343)
SES, γ_{010}	-4.979* (0.223)	-4.978* (0.224)	-4.983* (0.610)	-4.978* (0.607)
GRP-KOR*SES, γ_{011}	-2.241 (1.274)	-2.241 (1.275)	-2.237 (1.737)	-2.241 (1.738)
TCHRATE, γ_{020}	2.825* (0.545)	2.824* (0.546)	2.829* (1.145)	2.821* (1.139)
GRP-KOR*TCHRATE, γ_{021}	7.987* (3.383)	7.984* (3.385)	7.976* (6.086)	7.980* (6.090)
HW, γ_{100}	1.198* (0.006)	1.198* (0.006)	1.198* (0.010)	1.198* (0.010)
GRP_KOR*HW, γ_{101}	0.063* (0.026)	0.063* (0.026)	0.063 (0.046)	0.063 (0.046)
EXTR, γ_{200}	0.133* (0.002)	0.133* (0.002)	0.133* (0.004)	0.133* (0.004)
GRP_KOR*EXTR, γ_{201}	0.026* (0.011)	0.026* (0.011)	0.026* (0.015)	0.026* (0.015)
EBS, γ_{300}	0.500* (0.005)	0.500* (0.005)	0.500* (0.006)	0.500* (0.006)
GRP_KOR*EBS, γ_{301}	0.049* (0.022)	0.049* (0.022)	0.049 (0.036)	0.049 (0.036)
Random effect				
Intra-class correlation, ρ	0.040	0.040	0.040	0.040
Level-1 variance, σ^2	10.170	10.170	10.169	10.170
Level-2 variance, τ^2	0.427	0.428	0.428	0.429

* $p < .05$

Table 7. Parameter Estimates of the Conditional Model for English

Fixed effect coefficient (SE)	Software			
	SAS	R	Mplus	HLM
Intercept, γ_{000}	16.493* (0.085)	16.493* (0.085)	16.495* (0.181)	16.493* (0.180)
GRP_ENG, γ_{001}	-0.187 (0.106)	-0.187 (0.106)	-0.189 (0.212)	-0.187 (0.212)
SES, γ_{010}	-7.453* (0.394)	-7.451* (0.395)	-7.455* (0.667)	-7.450* (0.671)
GRP-ENG*SES, γ_{011}	3.497* (0.483)	3.496* (0.483)	3.497* (1.037)	3.495* (1.044)
TCHRATE, γ_{020}	3.757* (0.656)	3.754* (0.656)	3.755* (1.231)	3.751* (1.229)
GRP-ENG*TCHRATE, γ_{021}	-3.350* (1.257)	-3.347* (1.258)	-3.341 (2.678)	-3.345 (2.677)
HW, γ_{100}	1.222* (0.015)	1.222* (0.015)	1.222* (0.035)	1.222* (0.036)
GRP_ENG*HW, γ_{101}	-0.025 (0.016)	-0.025 (0.016)	-0.025 (0.036)	-0.025 (0.036)
EXTR, γ_{200}	0.108* (0.008)	0.107* (0.008)	0.107* (0.012)	0.107* (0.012)
GRP_ENG*EXTR, γ_{201}	0.030* (0.008)	0.030* (0.008)	0.030* (0.013)	0.030* (0.013)
EBS, γ_{300}	0.576* (0.015)	0.576* (0.015)	0.576* (0.022)	0.576* (0.022)
GRP_ENG*EBS, γ_{301}	-0.079* (0.016)	-0.079* (0.016)	-0.079* (0.023)	-0.079* (0.023)
Random effect				
Intra-class correlation, ρ	0.039	0.039	0.039	0.039
Level-1 variance, σ^2	10.170	10.170	10.169	10.170
Level-2 variance, τ^2	0.415	0.416	0.416	0.417

* $p < .05$

5. Discussion

The study fitting 2-level HLMs to the NAEA data showed that ability grouping influenced student achievement differently in the two subjects. The effects were positive in English and negative in Korean, respectively. The study also provided some evidence that ability grouping affected students' attitudes toward the curriculum. Looking at the significantly lower attitude rating about Korean in schools that had an ability-grouping policy, ability grouping may have caused negative side effects by lowering students' motivation and commitment for learning due to poor attitudes toward the subject. This finding was in line with previous studies which reported on the link between ability grouping and students' attitudes, self-efficacy, or self-esteem (Berends 1991; Oakes 2000; Sung Y. 2008; Vanfossen, Jones, & Spade 1987). For instance, ability grouping lowered self-efficacy, especially in the low ability group (Sung Y. 2008).

In contrast, attitude rating about English appeared lower for schools having an ability-grouping policy, but the difference between the two conditions was not significantly large. A previous study reported a compatible finding that 4th grade students showed little difference in self-esteem, motivation, and anxiety between two different ability groups in English (Shin G. 2006). Another study, however, showed that ability grouping improved college students' satisfaction about English classes (Kim Y. & Lee J. 2009; Kim H. & Lee H. 2010).

Students' background characteristics such as self-efficacy, self-reported understanding in class, academic commitment or delinquency, and so forth have been reported to be significantly associated with achievement (Ban J. & Shin S. 2011; Jenkins 1995; Kim S. & Kim J. 2011; Lee J. 1999). In the present study, ability grouping appeared to impact student attitude toward Korean negatively, and the average performance in the government-mandated Korean test was better when schools had no ability grouping policy for the subject. On the other hand, ability grouping did not appear to have a significantly negative impact on student attitude about English, and the schools having an ability-grouping policy for the subject performed better in the English test.

Many studies have reported that SES has a negative impact on student achievement (Harris & Sass 2007; Shin & Slater 2010). This

study showed that SES could also affect students' attitudes toward curriculum. The number of teachers relative to students also seemed to play a significant role in students' attitudes about curriculum. The lack of school resources including qualified teachers is known to be negatively related to student performance in the literature (Backhoff, Bouzas, Contreas, Hernández, & García 2007; Ban J. & Shin S. 2011). Students' efforts for after-school learning through doing homework or attending private tutoring or EBS online classes were reported to be positively related to academic success (Ban J. & Shin S. 2011). The findings of the present study showed that such efforts were also positively related to students' attitudes about the curriculum.

For checking multicollearity, correlation coefficients were computed. Overall, the predictors showed little correlations vis-à-vis the general rule for correlations (Stevens 2002). The correlations were .44 or lower. A general cut off for strong linear associations between variables is known to be ± .8 to ± .9 (Mason & Perreault 1991). It is well documented that the presence of strong correlations indicates collinearity. Collinearity may lead to inflated variances of some of the coefficients in the linear model and the structural equation model (Grewal, Cote, & Baumgartner 2004; Sengupta & Bhimasankaram 1997). The variance inflation is frequently measured by the Variance Inflation Factor (VIF). If the VIF is 10 or larger, the level of collinearity is believed to be problematic (Belsley, Kuh, & Welsch 1980). However, few statistical software packages provide the VIF in multilevel regression analysis. The authors could not find any option in any of SAS, R, Mplus, and HLM which produced this index. Further, the true effects of the VIF to hypothesis testing should be considered alongside other factors such as sample size, coefficient of determination, and so on (O'Brien 2007), as opposed to blindly deleting any variable associated with a VIF of 10 or larger.

All four software packages turned out to be quite comparable in that they yielded almost identical estimates of the fixed and random effects. The estimated standard errors associated with the effects, however, appeared different in some cases, e.g., GRP_KOR*HW (γ_{101}). Mplus and HLM produced larger values. Hence, the results of the significance test turned out different. The robust standard errors were presented here for the HLM case. It appears that parameter estimation methods and techniques are implemented in slightly different

ways across the software packages under investigation. Consequently, although the choice for the software to conduct a HLM analysis seems mostly up to researchers' preference on the basis of accessibility, familiarity, or convenience, if researchers are to be conservative in terms of Type I error, Mplus and HLM looked a better choice.

Last, it is worth noting that the current study used large-scale data. The sample size was huge. The statistical significance may be attributed primarily to the large sample size. In future research, the type of statistics such as effect size indices for multilevel data should be explored to judge the practical importance of the effect of each predictor in depth.

6. Conclusion

The current study explained not only the effects of ability grouping on students' achievement in Korean and English, but also their attitudes about the curricula. The study also explored student- and school-characteristics related to the students' attitudes. Since students' attitudes toward curricula must be linked directly and/or indirectly to their academic success, it seems important to comprehend the structure of various contextual factors surrounding the attitudes. Further, it seems that ability grouping is assumed to help schools promote academic excellence and educational equality simultaneously, considering that the vast majority of high schools in Korea have already employed it. Large-scale research such as the current study can support or challenge this assumption. The study showed that the effects of ability grouping on student academic achievement and student perception about curricula varied depending on which subject had implemented it. The study also reported that different computer software packages SAS, R, Mplus, HLM functioned quite comparably in estimating parameters, and that Mplus and HLM produced larger standard errors of the parameter estimates.

References

김소영, 김진한. (2011). 학교 특성이 중학생의 국어, 영어, 수학 성취도에 미치는 영향. 교육학연구 49.4, 205-229.

김연숙, 이지연. (2009). 대학 교양영어 수준별 수업에 대한 학습자와 교수자의 인식에 관한 연구. 영어교육, 64.4, 337-368.

김현옥, 이혜경. (2010). 대학 교양영어 수준별 프로그램 운영방안. 영어교육연구, 22.4, 75-101.

김현정. (2006). 수준별 이동수업에 대한 고등학교 학생들의 인식도 조사. 석사학위논문. 인하대학교 교육대학원.

반재천. (2006). 국가수준 학업성취도 평가의 척도개발. 교육평가연구 19.2, 113-135.

반재천, 신선희. (2011). 다수준 맥락변인이 학업성취도에 영향에 미치는 효과 추이 분석: 2003-2009년 국가수준 학업성취도 평가를 중심으로. 교육평가연구 24.3, 685-712.

서현경. (2008). 수학과 수준별 이동수업에 대한 학생들의 인식과 수준별 이동수업의 개선방안. 한국수학교육학회지 시리즈 E: 수학교육 논문집 22.3, 253-273.

성열관. (2008). 수준별 교육과정의 감환된 의미로서 영어, 수학 이동수업의 효과성 검토. 교육과정연구 26.2, 167-189.

신경원. (2006). 수준별 이동수업에 따른 교실영어가 학업성취도 및 정의적 태도에 미치는 영향. 석사학위논문, 중앙대학교.

이정원. (1999). 영어에 대한 자기조절 학습전략 사용과 자기효능감의 학업성취와의 관계. 교육과정평가연구 2.1, 169-184.

유준상. (2008). 수준별 이동수업이 성적변화에 미치는 영향에 대한 고등학생들의 인식 연구. 교육문화연구 14.1, 41-73.

최지나. (2006). 영어과 수준별 이동수업과 학업 성취도와의 관계. 석사학위논문. 숙명여자대학교 교육대학원.

황여정, 김경근. (2008). 수준별 이동수업의 교육효과에 대한 종단적 분석. 제2회 한국교육종단연구 학술대회 논문집, 327-349. 서울: 한국교육개발원.

Backhoff, E. Bouzas, A., Contreras, C., Hernández, E., & Gracia, M. (2007). Factores escolares y aprendizaje en México: El caso de la educación básica. México: Instituto Nacional para la Evaluación de la Educación.

Ball, S. J. (1981). *Beachside comprehensive: A case-study of secondary schooling*. Cambridge, England: Cambridge University Press.

Ban, J., & Shin, S. (2011). Trend of the effects of multilevel contextual variables on student achievement: 2003-2009 NAEA Data Analysis. *Korean Journal of Educational Evaluation* 24.3, 685-712.

Barr, R., & Dreeben, R. (1983). *How schools work*. Chicago, IL: University of Chicago Press

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York, NY: John

Wiley.

Berends, M. (1991). High school ability grouping and students' school orientations. Unpublished manuscript, University of Wisconsin, Madison.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage Publications.

Burgess, R.G. **(**1984). It's not a proper subject: It's just Newsom. In I. Goodson & S. Ball (Ed.), *Defining the Curriculum* (pp. 181-200). London, UK: The Falmer Press.

Foy, P., & Olson, J. F. (Eds.). (2009). *TIMSS 2007 technical report.*Chestnut Hill, MA: TIMSSInternational Study Center, Boston College.

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education* 60*,* 135-55.

Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review* 57, 812-828.

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Reinforcement, compensation, or neutrality? *American Journal of Sociology* 94, 1146-83.

Grewal, R., Cote, J. A., & Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: implications for theory testing. *Marketing Science* 23.4, 519-529.

Harris, D. N., & Sass, T. R. (2007). *The Effects of NBPTS-Certified Teachers on Student Achievement.* Retrieved from: http://www.teacherqualityresearch.org/effects.pdf

Hoffer, T. (1992). Middle school ability grouping and student achievement in science and mathematics. *Education Evaluation and Policy Analysis* 14*,* 205-27.

Jenkins, P. H. (1995). School delinquency and school commitment. *Sociology of Education* 68.3, 221-239.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). SAS system for mixed models. Cary, NC: SAS Institute.

Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing* 28, 268-280.

Mplus (Version 6.12). USA: Muthen & Muthen.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity* 41, 673-690.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

Oakes, J. (2000). *Keeping Track, Part 1: The Policy and Practice of Curriculum Inequality*. Retrieved from: http://equity.enc.org/equity/eqtyres/erg/111442/1442.htm.

OECD Programme for International Student Assessment. (2009). *PISA 2009 technical report*. Retrieved from http://www.pisa.oecd.org/

R (Version 2.15.0). Retrieved from http://cran.r-project.org/bin/windows/base/

Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking*. New York, NY: Wiley.

SAS (Version 9.2). USA: SAS Institute Inc.

Sengupta, D., & Bhimasankaram, P. (1997). On the roles of observations in collinearity in the linear model. *American Statistical Association* 92.439, 1024-1032.

Shin, S., & Slater, C. L. (2010). Principal leadership and mathematics achievement: An international comparative study. *School Leadership and Management* 30.4, 317-334.

Stevens, James P. (2002). *Applied Multivariate Statistics for the Social Sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Vanfossen, B. E., Jones, J. D., & Spade, J. Z. (1987). Curriculum tracking and status maintenance. *Sociology of Education* 60, 104222.

Verbeke, G., & Molenberghs, G. (2000). *Springer series in statistics: Linear mixed models for longitudinal data*. New York, NY: Springer-Verlag.

Appendix A. Research Variables and Corresponding Survey Questions

Variable	Type of variable	Level	Survey question
GRP_KOR	Predictor	2	Does your school employ ability grouping in the following subjects? 0 − No ability grouping in Korean 1 − Ability grouping in Korean
GRP_ENG	Predictor	2	Does your school employ ability grouping in the following subjects? 0 − No ability grouping in English 1 − Ability grouping in English
SES (P3/P2_4)	Predictor	2	The number of students from a welfare family: (P3) The total number of students: (P2_4)
TCHRATE (P14_1/P2_4)	Predictor	2	The number of teachers: (P14_1) The total number of students: (P2_4)
HW	Predictor	1	How many hours do you spend in doing homework assignments a day on average? 1 − None 2 − Less than 1 hour 3 − Between 1 -2 hours 4 − Between 2-3 hours 5 − More than 3 hours
EXTR(S18+S19)	Predictor	1	How many hours do you spend in private learning materials or private internet classes a day on average? (S18) How many hours do you spend in private tutors' or institutes' classes a day on average? (S19) 1 − None 2 − Less than 1 hour 3 − Between 1-2 hours 4 − Between 2-3 hours 5 − More than 3 hours
EBS	Predictor	1	How many hours do you spend in EBS or public internet classes a day on average? 1 − None 2 − Less than 1 hour 3 − Between 1-2 hours 4 − Between 2-3 hours 5 − More than 3 hours

Variable	Type of variable	Level	Survey question
YCURR (Sum of S7 through S14)	Criterion	1	I keep class materials well. (S7) I concentrate on lecture. (S8) I preview class contents. (S9) I review the contents covered in class. (S10) I ask my teacher questions regarding class contents in and right before or after the class. (S11) I actively involves in experiment, discussion, group activities in class. (S12) I can understand the subject content from learning in class. (S13) I can study by myself without attending private tutors'or institutes' aids. (S14) 1 − Strongly disagree 2 − Disagree 3 − Agree 4 − Strongly agree
KOR-AL	Criterion	1	Performance level 1 − Advanced 2 − Proficient 3 − Basic 4 − Below basic

Seon-Hi Shin, Ph.D.
Measurement and Statistical Research Department
TEPS center, Language Education Institute
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea
Email: shinseonhi@snu.ac.kr

Seohyun Kim, M.S.
Statistics Department
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea

Received : October 31, 2012
Revised version received: December 12, 2012
Accepted: December 22, 2012